This document describes the construction of the multinational production data set used together with the project.

**Sources of data:**

1. Romondo, Rodriguez-Clare and Tintelnot (2015): 61 countries, 1996-2001 average
2. Eurostat
3. UNCTAD bilateral FDI stocks and flows, 2001-2012
4. OECD

**Description of the preparation of each dataset**

1. Offered by **Romondo** from earlier email exchanges.
2. **Eurostat** (see do-file: prepareData/eurostat\_fdi.do)

The do file starts from the “fats” tables downloaded from Eurostat.

To construct aggregate multinational sales, the following tables and variables are used

|  |  |  |  |
| --- | --- | --- | --- |
| Table | Variables | Industry | Years |
| fats\_g1b\_08: total inward MP activities | V12110: Turnover or gross premiums written | Nace Rev2 B-N\_S95\_X\_K: Foreign control of enterprises by controlling countries (from 2008 onwards) NACE Rev 2 | 2008-2012 |
| fats\_g1b\_03: total inward MP activities | V12110 | Nace Rev1 C-K\_X\_J: Business economy - Industry and services (except financial intermediation) | 2003-2007 |
| fats\_96/fats\_sum | V12110 | NACE Rev. 1.1 C-K\_X\_J | 1999-2002 |
| fats\_out1: outward MP activities by broad industries | TUR: Turnover - Million ECU/EUR | Nace Rev1 A-O\_X\_L: All NACE activities (except public administration; activities of households and extra-territorial organizations) | 2004-2006 |
| fats\_out2: outward MP activities by broad industries | TUR | Nace Rev1 A-O\_X\_L | 2007-2009 |
| fats\_out2\_r2 | TUR | Nace Rev2 B-S\_X\_O: Industry, construction and services (except public administration, defense, compulsory social security) | 2010-2012 |

For years before 2002, only 11 countries voluntarily report bilateral FATS tables. Among them, Ireland and Germany do not report total values (sector C-K\_X\_J). I aggregate industry level values to get the total. For Ireland, I aggregate NACE Rev1.1 sectors “C D E G H I K”, requiring no missing values in any of the sectors.

To construct aggregate flows and stocks, use the following tables:

|  |  |  |  |
| --- | --- | --- | --- |
| Tables | Description |  | Year |
| tec00049 | Direct investment inward flows by main investing country | EU inflow/outflow with rest of the world | 2001-2013 |
| tec00051 | Direct investment inward stocks by main origin of investment |
| tec00053 | Direct investment outward flows by main destinations |
| tec00052 | Direct investment outward stocks by main destinations |
| bop\_fdi\_pos\_r2, bop\_fdi\_flow\_r2 | EU direct investment positions (flows), breakdown by country and economic activity (NACE Rev. 2) | bilateral FDI flows and stocks | 2008-2012 |
| bop\_fdi\_pos, bop\_fdi\_flows | NACE Rev. 1 | bilateral FDI flows and stocks | 1980-2009 |

For years 2008-2009, both NACE Rev1 and Rev2 data exist. Since for most of the matched country pairs, the total FDI values from the two versions are exactly the same (see Output/check\_data/tables/eurostat\_bop\_vs\_tec.csv), I use Rev2 data whenever possible, and update missing values with Rev1 data. For the year 2008-2009, about 15,000 values are from Rev2 and a few hundred missing values are updated.

TEC tables contain data on main origins and destinations from 2001 to 2013. Country pairs matched with BOP tables have exactly the same values. Thus I use BOP tables whenever possible. Updating missing values using TEC helps me to complete some of the values for 2013.

1. **UNCTAD data**

UNCTAD data can be downloaded from their [website](http://unctad.org/en/Pages/DIAE/FDI%20Statistics/FDI-Statistics-Bilateral.aspx). They are collected from national sources and international organizations. The downloaded data is one excel for each country. After consolidating the data, I perform additional cleaning (see code Code/prepareData/UNCTAD\_bilateral.do and it is companion record Output/tables/clean\_n\_checking/UNCTAD\_bilateral.txt).

* 1. In a few cases, countries report FDI flows from itself for some reason and I simply dropped these cases.
  2. Because both origin and destination countries’ tables may contain numbers reported by one of the two countries, there are duplicate items in terms of origin, destination and reporting country in each year. In very few cases the two numbers are both non-zero or nonmissing. If one is missing and the other is not, I use the nonmissing value. If one is zero and the other is not, I use the non-zero value. If both are nonmissing and nonzero, I use their average if the absolute percentage difference is less than 10%.
  3. Among the 80,000 home-host-year combinations with nonmissing FDI flows, 20,000 are reported both by home and host. 40,000 only have host country reporting while 20,000 only have home country reporting. Similar numbers for the FDI stocks. The numbers reported by the host and home countries differ greatly. (see Code/checkData/checkFDI\_sources.do)

1. **OECD Data**

OECD also compiles data on FDI and multinational activities, based on statistics reported by OECD countries. The multinational firm activity data can be downloaded from OECD database – Globalization – Activity of Multinationals. There are 12 tables in this section

|  |  |  |
| --- | --- | --- |
| Tables | Description | Year |
| Inward activity of multinationals by industrial sector - ISIC Rev 4 | Total multinational activities from all other countries, 79 ISIC Rev 4 sectors, 17 variables (turnover and employment are best covered) | 2008-2013 |
| Inward activity of multinationals by investing country - ISIC Rev 4 | 17 variables, 200ish partner countries, 5 main sectors, 17 variables | 2008-2013 |
| Outward activity of multinationals by industrial sector - ISIC Rev 4 | Total multinational activities from all other countries, 14 variables, 79 ISIC Rev 4 sectors | 2007-2013 |
| Outward activity of multinationals by country of location - ISIC Rev 4 | 14 variables, 3 main industries (**cannot exclude finance**), 200ish partner countries | 2007-2013 |
| Inward activity of multinationals by industrial sector (manufacturing) - ISIC Rev 3 | Total multinational activities from all other countries, 16 variables, 71 ISIC Rev 3 sectors (including manufacturing and services), a few countries have two sources | 1985-2013 |
| Inward activity of multinationals by investing country, total manufacturing - ISIC Rev 3 | Total manufacturing only | 1985-2013 |
| Inward activity of multinationals - Share in national total (manufacturing) |  | 1985-2011 |
| Outward activity of multinationals by industrial sector (manufacturing) - ISIC Rev 3 | Total multinational activities from all other countries, 15 variables, 70 ISIC Rev 3 sectors, 1985-2013 | 1985-2013 |
| Outward activity of multinationals by country of location - ISIC Rev 3 | Manufacturing and total business, 69 partner countries | 1985-2013 |
| Outward activity of multinationals - Share in national total (manufacturing) |  | 1989-2009 |
| Inward activity of Multinationals in ISIC Rev 3 (services) | 12 variables, 200ish partner countries, 81 sectors (including manufacturing) | 1995-2008 |
| Outward Activity of Multinationals in ISIC Rev 3 (services) | 12 variables, 200ish partner countries, 83 sectors (including manufacturing) | 1995-2009 |

To construct a measure of bilateral multinational sales, I use the four highlighted data sets. The definitions for “total activities” are slightly different across them.

|  |  |
| --- | --- |
| Inward activity of multinationals by investing country - ISIC Rev 4 | TOTAL ACTIVITY (sec B to N excl. K) |
| Outward activity of multinationals by country of location - ISIC Rev 4 | TOTAL BUSINESS SECTOR (sec B to S excl. O[[1]](#footnote-1)) |
| Inward activity of Multinationals in ISIC Rev 3 (services) | GRAND TOTAL (01-93 minus 75[[2]](#footnote-2)); for some country pairs, data on Financial intermediation (65-67) is available |
| Outward Activity of Multinationals in ISIC Rev 3 (services) | GRAND TOTAL (01-93 minus 75); for some country pairs, data on Financial intermediation (65-67) is available |

For years before 2007 (including 2007), ISIC Rev 3 data provide most of the information, while for years after 2007, ISIC Rev 4 data provide most of the information. Thus I ignore ISIC Rev 3 data for 2008 and 2009, and ignore ISIC Rev 4 data for 2007. For numbers of common country pairs, see “Output/tables/clean\_n\_checking/oecd\_fdi.txt”.

Note that the monetary variables are in millions of local currency. According to the meta data in the online database,

For Euro area countries, national currency data is expressed in euro beginning with the year of entry into the Economic and Monetary Union (EMU). For years prior to the year of entry into EMU, data have been converted from the former national currency using the appropriate irrevocable conversion rate. This presentation facilitates comparisons within a country over time and ensures that the historical evolution is preserved. Please note, however, that pre-EMU euro are a notional unit and should not be used to form area aggregates or to carry out cross-country comparisons.

German data on outward MNE sales are outliers in 2007. It seems the undocumented table “Outward Activity of Multinationals in ISIC Rev 3 (services)” present values that are 1000 times larger. For example, in the OECD database, table “Outward Activity of multinationals by industrial sector (manufacturing) – ISIC Rev 3” shows that Germany’s total outward sales (Total Business Enterprise) is 1.497 million Euros in 2007. This number, however, is 1.486 billion Euros in the first table. Thus, I rescale all values for Germany in 2007 by 1000 times.

Slovenia presents another anomaly in the data. The grand total outward sales to the WORLD are 3394715 million in local currency in 2007, and there seems to be no break after that (Slovenia adopted Euro in 2007, and the table only contains data for 2007-2009). I suspect that they are still denominated in Slovenian Tolar rather than Euros. Thus for these three years, I transformed the Euro exchange rate to notional Tolar exchange rate using the fixed exchange rate for Solvenia.

1. **Number of employees and Number of persons employed**

In OECD and Eurostat inward tables, there are two variables related to total employment of multinational firms: (1) number of employees and (2) number of persons employed. A detailed definition of these two concepts can be found on Page 56-57 in the Eurostat Foreign AffiliaTes Statistics (FATS) Recommendations Manual (2012 edition). The manual does not provide a direct comparison between the two concepts but the key difference between the two concepts seems to be that one is counted as an employee only when a contract of employment is provided.

For country-pair-years which both variables are nonmissing, I can calculate the difference between the two variables. Most of them are exactly the same, and (2) is in general larger than (1). When possible, I use (2) as a measure of employment and supplement (1) if (2) is missing or zero but (1) is positive.

1. **Compare MNE activities from OECD and Eurostat**

|  |  |  |
| --- | --- | --- |
|  | OECD | Eurostat |
| Inward (year<=2007) | Include financial sectors | Exclude financial sectors |
| Inward (year>=2008) | Exclude financial sectors | Exclude financial sectors |
| Outward | include financial sectors | include financial sectors |

First, I need to exclude the financial sector for all the bilateral multinational sales. Both OECD and Eurostat recommend that value of characteristics should be allocated to the main activity of the foreign affiliate. (see OECD Benchmark Definition of FDI 4e, p115; Eurostat 2012, FATS Recommendations Manual, p33) Therefore, if we know the sales breakdown of the financial sector, we can subtract those from sales of the Total Business Sector.

For inward sales (OECD before 2007), I first subtract sales of the financial sector (ISIC Rev 3 65-67) if it is available. This helps me to correct inward sales for xxx out of xxxxx observations. For the rest of the inward total sales, I assume the fraction that can be attributed to the financial sector is the same across origin countries and use the fraction of financial sales from the “WORLD” to adjust for the values if available.[[3]](#footnote-3) This helps me correct xxx observations. For the rest of the observations, I use the host countries’ share of financial sector output. If the host countries’ share of financial sector output is not available, I use the home countries’ share of financial sector output. The same procedure is applied to OECD outward sales and the only difference is that I use nonfinancial share of total outward MP in Step 2.[[4]](#footnote-4) However, since OECD database does not have bilateral MNE sales by industry after 2007, I only perform the second and third steps.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | OECD inward before 2007 | OECD outward | Eurostat outward |
| Bilateral sales needs to be adjusted |  | 5762 | 12714 | 7614 |
| Adjusted Using: | Bilateral financial sales | 1970 | 3092 | 3260 |
|  | Total outward financial sales share | 2040 | 9162 | 3354 |
|  | Nonfinancial output share of the destination (three year average) | 1752 | 191 | 51 |
|  | Nonfinancial output share of the destination (2010-2012, using last year available) |  | 269 | 343 |
| Not adjusted |  | 0 | 0 | 606 |

1. **Consolidate three sources of FDI statistics**

Given MNE activity data from OECD and Eurostat, and FDI stocks and flows data from OECD, Eurostat and UNCTAD, I try to impute some of the missing bilateral MNE activity variables (mainly revenue, but employment is also considered). I use the following steps

* + 1. Combine information in number of employees (n\_emp) and number of persons employed (n\_psn\_emp). Use number of persons employed as the primary source and supplement with number of employees. See the do file code/check\_data/emp\_vs\_psn\_emp.do and the companion output for the comparison between the two variables. Most of the numbers are the same, but on average, number of employees are smaller than number of persons employed since the former only include workers with employment contract.
    2. Drop outliers in bilateral relationships defined using year-to-year growth rates

I look for outliers within a home-host country pair. I compute both the deviation from the log mean and the log change in a certain variable, and define an observation to be an outlier if it satisfies two conditions (1) the log change from last period > 5 or < -5, or the log change into next period >5 or < -5; (2) the deviation from the log mean > 5 or < -5. Note that in this way, we first take care of the zeros since they won’t enter the candidates for outliers. Second, the observation adjacent to the outlier is not likely to be misidentified as an outlier since it will be close to the mean. There are a few scenarios in which I might fail to identify an outlier: (1) if an outlier has no adjacent observations (2) if the value of the outlier is large enough to make the average close to itself, so it does not satisfy condition (2) and will not be identified as outlier. However, the adjacent values might be identified as an outlier since they are likely to be away from the mean. In the tables in “output/data\_management/tables/potential\_outliers.xlsx”, I list all the country pairs which have at least one observation satisfying condition (2).

* + 1. Supplement missing data in Eurostat with OECD – document them well

First, identify countries that report Eurostat or OECD. If a country reports in Eurostat, use Eurostat as the primary source. If a country reports in OECD, use OECD as the primary source.

* + 1. Impute additional zeros using FDI stocks

If one of the Eurostat or OECD stock is non-positive, and the other is missing or non-positive, impute the missing MNE activities (employment or revenue) to be zero. If employment is zero, impute revenue to be zero too (very few observations).

* + 1. Extrapolate still missing sales data with inward FDI stocks (and outward revenue) using **both cross-sectional and over time variation.** For the period 1995 – 2012, I run the following regression for host and home countries with at least three observations

,

where the dependent variable is the inward sales while the key independent variable is either inward stock or outward sales. I estimate this regression using different time periods and the coefficient b seems very stable. (see output/data\_management/extrap\_activities.csv)

* + 1. Next I consider using the time-series property of the data only and do not bring in any new variable besides inward sales. I extrapolate over time using a constant growth model within each pair. For country pairs with at least 4 observations between 2001 and 2012, I estimate the following equation

Note that I do not impose a pair-specific trend since the trend can be imprecisely estimated with only a few observations within a pair. Instead, I impose the growth rates to have host and home specific components, and also a global trend, which is not restricted to be linear.

Besides the linear growth model extrapolation for observations with positive sales, I also impute additional zeros using the time series data. I first identify consecutive runs of missing values or zeros. I require the missing values can potentially be replaced as zeros, i.e., it cannot have positive stock and other MNE activities (inward employment, number of enterprises and outward employment, sales and number of enterprises). After such runs are identified, I identify missing values that are “squeezed” between two zeros. These values are replaced with zeros.

1. **Extrapolation for total inward activities**

The extrapolation for total inward activities is a bit easier. The procedures are similar to the extrapolation of bilateral activities. I describe the procedures as follows.

* + 1. Combine information in number of employees (n\_emp) and number of persons employed (n\_psn\_emp). Use number of persons employed as the primary source and supplement with number of employees.
    2. Drop outliers in total inward/outward variables defined using year-to-year growth rates

I compute both the deviation from the log mean and the log change in a certain variable, and define an observation to be an outlier if it satisfies two conditions (1) the log change from last period > 5 or < -5, or the log change into next period >5 or < -5; (2) the deviation from the log mean > 5 or < -5.

* + 1. Supplement missing data in Eurostat with OECD – document them well

First, identify countries that report Eurostat or OECD. If a country reports in Eurostat, use Eurostat as the primary source. If a country reports in OECD, use OECD as the primary source.

* + 1. Impute additional zeros using FDI stocks

If one of the Eurostat or OECD stock is non-positive, and the other is missing or non-positive, impute the missing MNE activities (employment or revenue) to be zero. If employment is zero, impute revenue to be zero too (very few observations).

* + 1. Extrapolate still missing sales data with inward FDI stocks and employment using **both cross-sectional and over time variation.** For the period 1995 – 2012, I run the following regression for host and home countries with at least three observations

,

where the dependent variable is the inward sales while the key independent variable is either inward stock or employment. I estimate this regression using different time periods and the coefficient b seems very stable. (see output/data\_management/extrap\_tot\_in\_activities.csv)

* + 1. Next I consider using the time-series property of the data only and do not bring in any new variable besides inward sales. I extrapolate over time using a constant growth model within each pair. For country pairs with at least 6 observations between 2001 and 2012, I estimate the following equation

Besides the linear growth model extrapolation for observations with positive sales, I also impute additional zeros using the time series data. I first identify consecutive runs of missing values or zeros. I require the missing values can potentially be replaced as zeros, i.e., it cannot have positive stock and other MNE activities (inward employment, number of enterprises and outward employment, sales and number of enterprises). After such runs are identified, I identify missing values that are “squeezed” between two zeros. These values are replaced with zeros.

1. **Inward v.s. Outward.**

Ramondo, Rodriguez-Clare and Tintelnot (2015) give two reasons for using outward sales as the primary source. First, they argue statistics reported by ­­the host country is more likely on “immediate owners” rather than “ultimate beneficiary owners (UBO)”. Second, sales reported by the host country may be only for local sales and miss sales from all other countries.

This may be true for their UNCTAD data (though I highly doubt since some of the UNCTAD data should come from Eurostat and OECD). However, I cannot find support for their arguments in FATS statistical manuals. For the first argument, I found the following related points in the relevant manuals

|  |  |  |
| --- | --- | --- |
| Source | Inward FATS | Outward FATS |
| UNCTAD Manual | As far as possible, it is recommended that countries use the UBO unit when compiling operational statistics for inward investment (Volume 2, II.39)[[5]](#footnote-5) | The second issue deals with the treatment of foreign investments of those domestic enterprises, which are themselves foreign-owned. This volume recommends that the compiling country should collect data for all resident enterprises direct investor, regardless of where they are owned. However, in its published statistics it should provide separate breakdowns for the foreign affiliates of domestically and foreign-owned enterprises. (Volume 2, II.41) |
| FATS Manual (Eurostat) | Ultimate Control Institutional Unit (UCI) is recommended (see I.1.1) | Ultimate Control Institutional Unit (UCI) is recommended (see I.1.1) |

UNCTAD recommends that countries use the UBO unit when compiling operational statistics for inward investment (activities), but IMF does require that BOP statistics record transactions based on the immediate foreign owner (II.39, II.40 in UNCTAD, 2008, Vol 2). On the contrary, UNCTAD recommends countries report outward MNE activities based on immediate owners (II.41 (ii)). I simply cannot find any information about their second argument.

1. Public administration and defence; compulsory social security [↑](#footnote-ref-1)
2. Public administration and defence; compulsory social security [↑](#footnote-ref-2)
3. I use a three year average whenever possible, ignoring missing years. If it is the first year (1995) or last year (2007) in the data, I use that year only. If [↑](#footnote-ref-3)
4. Again I use a three year average. Now even for the first and last years in the data I can obtain a three year average. However, there is no data on financial output after 2009 for most countries. For the later years (2010-2012) I use the non-financial output share of that country in the **last year available**. [↑](#footnote-ref-4)
5. The footnote in that paragraph reads: Out of 15 countries providing operational data in the OECD’s Manual (OECD, 2001) eight (Belgium, France, Germany, Japan, Luxembourg, Norway, Poland and Portugal) use immediate foreign owner, and seven (Finland, Ireland, Italy, the Netherlands, Sweden, the United Kingdom and the United States) use ultimate owner. [↑](#footnote-ref-5)